

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

IN RE APPLICATION OF: Katsuji WATANABE, et al.

GAU:

SERIAL NO: New Application

EXAMINER:

FILED: Herewith

FOR: METHOD AND SYSTEM FOR SEARCHING FOR RELATIONSHIPS BETWEEN BASE SEQUENCES IN GENES



REQUEST FOR PRIORITY

ASSISTANT COMMISSIONER FOR PATENTS
WASHINGTON, D.C. 20231

SIR:

- ☐ Full benefit of the filing date of U.S. Application Serial Number, filed, is claimed pursuant to the provisions of 35 U.S.C. §120.
- ☐ Full benefit of the filing date of U.S. Provisional Application Serial Number, filed, is claimed pursuant to the provisions of 35 U.S.C. §119(e).
- ☒ Applicants claim any right to priority from any earlier filed applications to which they may be entitled pursuant to the provisions of 35 U.S.C. §119, as noted below.

In the matter of the above-identified application for patent, notice is hereby given that the applicants claim as priority:


<u>COUNTRY</u>	<u>APPLICATION NUMBER</u>	<u>MONTH/DAY/YEAR</u>
JAPAN	2000-215134	July 14, 2000

Certified copies of the corresponding Convention Application(s)

- ☒ are submitted herewith
- ☐ will be submitted prior to payment of the Final Fee
- ☐ were filed in prior application Serial No. filed
- ☐ were submitted to the International Bureau in PCT Application Number .
Receipt of the certified copies by the International Bureau in a timely manner under PCT Rule 17.1(a) has been acknowledged as evidenced by the attached PCT/IB/304.
- ☐ (A) Application Serial No.(s) were filed in prior application Serial No. filed ; and
(B) Application Serial No.(s)
 - ☐ are submitted herewith
 - ☐ will be submitted prior to payment of the Final Fee

Respectfully Submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.



Norman F. Oblon
Registration No. 24,618
C. Irvin McClelland
Registration Number 21,124



22850

OSP-10605 45

日 本 国 特 許 庁
PATENT OFFICE
JAPANESE GOVERNMENT



別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日

Date of Application:

2000年 7月14日

出 願 番 号

Application Number:

特願2000-215134

出 願 人

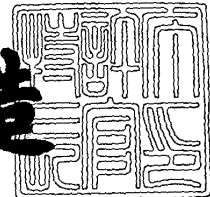
Applicant(s):

農林水産省九州農業試験場長
渡邊 克二

2001年 3月 9日

特許庁長官
Commissioner,
Patent Office

及 川 耕 造



出証番号 出証特2001-3018010

【書類名】 特許願

【整理番号】 J84516A1

【提出日】 平成12年 7月14日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/15
C12N 15/00

【発明の名称】 遺伝子の類縁性検索方法および遺伝子の類縁性検索システム

【請求項の数】 15

【発明者】

 【住所又は居所】 熊本県菊池郡西合志町大字須屋 2 4 2 1 農林水産省九州農業試験場内

 【氏名】 渡邊 克二

【発明者】

 【住所又は居所】 熊本県菊池郡西合志町大字須屋 2 4 2 1 農林水産省九州農業試験場内

 【氏名】 奥田 充

【特許出願人】

 【持分】 099/100

 【識別番号】 591224434

 【氏名又は名称】 農林水産省九州農業試験場長

【特許出願人】

 【持分】 001/100

 【住所又は居所】 熊本県菊池郡西合志町大字須屋 2 4 2 1 農林水産省九州農業試験場内

 【氏名又は名称】 渡邊 克二

【代理人】

 【識別番号】 100064908

 【弁理士】

【氏名又は名称】 志賀 正武

【選任した代理人】

【識別番号】 100108578

【弁理士】

【氏名又は名称】 高橋 詔男

【選任した代理人】

【識別番号】 100089037

【弁理士】

【氏名又は名称】 渡邊 隆

【選任した代理人】

【識別番号】 100101465

【弁理士】

【氏名又は名称】 青山 正和

【選任した代理人】

【識別番号】 100094400

【弁理士】

【氏名又は名称】 鈴木 三義

【選任した代理人】

【識別番号】 100107836

【弁理士】

【氏名又は名称】 西 和哉

【選任した代理人】

【識別番号】 100108453

【弁理士】

【氏名又は名称】 村山 靖彦

【手数料の表示】

【予納台帳番号】 008707

【納付金額】 210円

【その他】 国以外のすべての者の持分の割合 1 / 1 0 0

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【物件名】 持分契約書 1

【提出物件の特記事項】 追って補充する

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 遺伝子の類縁性検索方法および遺伝子の類縁性検索システム

【特許請求の範囲】

【請求項 1】 制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データとして出力する理論値計算部と、

前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較部と

を備えることを特徴とする塩基配列類縁性検索システム。

【請求項 2】 前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示部を備えることを特徴とする請求項 1 に記載の塩基配列類縁性検索システム。

【請求項 3】 前記表示部は、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする請求項 2 に記載の塩基配列類縁性検索システム。

【請求項 4】 前記比較部は、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とする請求項 1 から 3 までのいずれかに記載の塩基配列類縁性検索システム。

【請求項 5】 プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の

増幅後塩基配列データを生成する増幅配列認識部を備え、

前記理論値計算部は、この増幅後塩基配列データを基に前記切断長理論値を計算する

ことを特徴とする請求項 1 から 4 までのいずれかに記載の塩基配列類縁性検索システム。

【請求項 6】 制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データを出力する理論値計算過程と、

前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較過程と

を有することを特徴とする塩基配列類縁性検索方法。

【請求項 7】 前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示過程を有することを特徴とする請求項 6 に記載の塩基配列類縁性検索方法。

【請求項 8】 前記表示過程では、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする請求項 7 に記載の塩基配列類縁性検索方法。

【請求項 9】 前記比較過程では、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とする請求項 6 から 8 までのいずれかに記載の塩基配列類縁性検索方法。

【請求項 10】 プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パタ

ーンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データを生成する増幅配列認識過程を有し、

前記理論値計算過程では、この増幅後塩基配列データを基に前記切断長理論値を計算する

ことを特徴とする請求項 6 から 9 までのいずれかに記載の塩基配列類縁性検索方法。

【請求項 1 1】 制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データを出力する理論値計算過程と、

前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較過程と

の処理をコンピュータに実行させるコンピュータプログラムを記録したコンピュータ読取り可能な記録媒体。

【請求項 1 2】 前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示過程の処理をコンピュータに実行させるコンピュータプログラムを記録した請求項 1 1 に記載のコンピュータ読取り可能な記録媒体。

【請求項 1 3】 前記表示過程では、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする請求項 1 2 に記載のコンピュータ読取り可能な記録媒体。

【請求項 1 4】 前記比較過程では、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とする請求項 1 1 から 1 3 までのいずれかに記載のコンピュータ読取

り可能な記録媒体。

【請求項 1 5】 プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データを生成する増幅配列認識過程の処理をコンピュータに実行させるコンピュータプログラムが記録されており、

前記理論値計算過程では、この増幅後塩基配列データを基に前記切断長理論値を計算する

ことを特徴とする請求項 1 1 から 1 4 までのいずれかに記載のコンピュータ読取り可能な記録媒体。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

この発明は、塩基配列類縁性検索方法および塩基配列類縁性検索システムに関する。特に、制限酵素による切断片の切断片長データを基にした塩基配列類縁性検索方法および塩基配列類縁性検索システムに関する。

【0 0 0 2】

【従来の技術】

試料として与えられた DNA (deoxyribonucleic acid, デオキシリボ核酸) 分子内の塩基配列を同定するために、その試料を分析することによって全塩基配列を出力するシーケンサ装置が広く用いられている。

【0 0 0 3】

また、制限酵素を用いて DNA 分子を切断し、切断片の長さを基に塩基配列を推定する方法もある。

【0 0 0 4】

【発明が解決しようとする課題】

しかしながら、上記シーケンサ装置は高価であり、より安価に塩基配列を特定できる手段が求められている。

【0 0 0 5】

また、制限酵素による切断片の長さを基に塩基配列を推定する方法は、様々な制限酵素によって得られる長さデータを用いて、既知の塩基配列とのパターンマッチングを行う手間が膨大であり、効率的なデータ処理が課題となっている。

【0006】

本発明は、上記事情に鑑みてなされたものであり、制限酵素による切断片長の実測データを基に、効率的に既知の塩基配列との類縁性を解析するための塩基配列類縁性検索方法および塩基配列類縁性検索システムを提供することを目的とする。

【0007】

【課題を解決するための手段】

上記の課題を解決するために、本発明による塩基配列類縁性検索システムは、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データとして出力する理論値計算部と、前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較部とを備えることを特徴とする。

【0008】

また、本発明による塩基配列類縁性検索システムは、前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示部を備えることを特徴とする。

【0009】

また、本発明による塩基配列類縁性検索システムは、前記表示部が、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料

とを関連付けて表示することを特徴とする。

【 0 0 1 0 】

また、本発明による塩基配列類縁性検索システムは、前記比較部が、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とする。

【 0 0 1 1 】

また、本発明による塩基配列類縁性検索システムは、プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データを生成する増幅配列認識部を備え、前記理論値計算部は、この増幅後塩基配列データを基に前記切断長理論値を計算することを特徴とする。

なお、ここで「増幅」とは、所定の上流プライマあるいは下流プライマまたはこれら両者を用いて、塩基配列の、始端から当該上流プライマ固有の塩基配列パターンの箇所までと当該下流プライマ固有の塩基配列パターンの箇所から終端までを切断し、残りの塩基配列の部分のみを抽出することを言う。

【 0 0 1 2 】

また、本発明による塩基配列類縁性検索方法は、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データを出力する理論値計算過程と、前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較過程とを有することを特徴とする。

【 0 0 1 3 】

また、本発明による塩基配列類縁性検索方法は、前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示過程を有することを特徴とする。

【 0 0 1 4 】

また、本発明による塩基配列類縁性検索方法は、前記表示過程では、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする。

【 0 0 1 5 】

また、本発明による塩基配列類縁性検索方法は、前記比較過程では、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とする。

【 0 0 1 6 】

また、本発明による塩基配列類縁性検索方法は、プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データを生成する増幅配列認識過程を有し、前記理論値計算過程では、この増幅後塩基配列データを基に前記切断長理論値を計算することを特徴とする。

【 0 0 1 7 】

また、本発明は、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データを出力する理論値計算過程と、前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長

実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較過程との処理をコンピュータに実行させるコンピュータプログラムを記録したコンピュータ読取り可能な記録媒体を要旨とする。

【0018】

また、本発明によるコンピュータ読取り可能な記録媒体は、前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示過程の処理をコンピュータに実行させるコンピュータプログラムを記録したことを特徴とする。

【0019】

また、本発明によるコンピュータ読取り可能な記録媒体は、前記表示過程では、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とするものである。

【0020】

また、本発明によるコンピュータ読取り可能な記録媒体は、前記比較過程では、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とするものである。

【0021】

また、本発明によるコンピュータ読取り可能な記録媒体は、プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データを生成する増幅配列認識過程の処理をコンピュータに実行させるコンピュータプログラムが記録されており、前記理論値計算過程では、この増幅後塩基配列データを基に前記切断長理論値を計算することを特徴とするものである。

【0022】

【発明の実施の形態】

以下、図面を参照しこの発明の一実施形態について説明する。

【0023】

図 1 は、この発明の一実施形態による塩基配列解析システム（遺伝子の類縁性検索システム）の構成を示すブロック図である。図 1 において、符号 2 1 は、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データである。また、2 2 b は、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている増幅後塩基配列データ（塩基配列データ）である。なお、この増幅後塩基配列データ 2 2 b は、一般に公開されている遺伝子データを手作業等の何らかの方法によって編集することにより得られる。

なお、増幅前の塩基配列データを基に自動的に増幅後塩基配列データ 2 2 b を生成する方法は、後の第二実施形態において説明する。

【 0 0 2 4 】

1 1 は、制限酵素データ 2 1 と増幅後塩基配列データ 2 2 b とを基に既知遺伝子を制限酵素で切断したときの切断長理論値を計算して、切断長理論値データ 2 3 を出力する理論値計算部である。理論値計算部 1 1 によって出力される切断長理論値データ 2 3 では、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持されている。また、1 3 は、この切断長理論値データ 2 3 と、実測値入力部 1 2 から入力された切断長実測値データ 2 5 とを比較することにより既知遺伝子と試料との類似度を算出して解析結果データ 2 7 として出力する比較部である。なお、この切断長実測値データ 2 5 では、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている。

【 0 0 2 5 】

1 4 はデータの比較の際の許容誤差を設定する許容誤差設定部であり、許容誤差設定部によって設定された値は許容誤差データ 2 9 として保持されている。比較部 1 3 は、この許容誤差データ 2 9 を参照して、切断長理論値データ 2 3 と切断長実測値データ 2 5 とを比較する際には、許容誤差範囲内の相違は一致しているものとして扱う。なお、比較部 1 3 による具体的な比較方法および許容誤差の適用方法については後に詳述する。

【 0 0 2 6 】

そして、15は、比較部13によって出力された解析結果データ27を基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示部である。表示部15による具体的な表示方法については、後で説明する。

【0027】

次に、切断長理論値の計算の基となる増幅後塩基配列データ22bの詳細について説明する。図2は増幅後塩基配列データ22b（塩基配列データ）の前半部分を示す図であり、配列番号1は同データの後半部分を示す図である（図3参照）。これらの図に示すように増幅後塩基配列データ22bはテキスト形式のデータであり、図2に示す前半部分には、下線を付したDNA番号（ACCESSION）と細菌名（ORGANISM）と遺伝子名（KEYWORDS）とが少なくとも含まれている。また、図3の後半部分には、塩基配列（ORIGIN）が含まれている。この塩基配列のデータ内の文字「a」はアデニン（adenine）を、「g」はグアニン（guanine）を、「c」はシトシン（cytosine）を、「t」はチミン（thymine）をそれぞれ表している。

【0028】

次に制限酵素データ21の詳細について説明する。図4は、制限酵素データ21のデータ構造およびデータ例を示す表図である。図4に示すように、制限酵素データ21は、表形式のデータであり、制限酵素番号と制限酵素名と切断箇所と切断位置の列を持っている。制限酵素番号は、制限酵素の識別のために本システムによって与えられた番号である。また、切断箇所は制限酵素が切断する箇所の塩基配列パターンを表し、切断位置はこの塩基配列パターン内における切断位置の変位を表す。例えば、図4に示す制限酵素番号12のSmaIは、DNA分子の塩基配列中の「・・・cccggg・・・」にマッチする箇所の変位3の位置に作用して、「・・・ccc」と「ggg・・・」に切断する。

【0029】

図5は、理論値計算部11に対して切断長理論値計算を指示する利用者インタフェース例を示す画面図である。この画面内には、存在している塩基配列ファイル（増幅後塩基配列データ22b）と制限酵素データ21内の制限酵素名とがそ

れぞれリスト表示されている。利用者がこれらのリストの中からそれぞれ選択して画面下部の「切断長計算」ボタンを押すことにより、理論値計算部 1 1 は選択されたデータを基に切断片長の理論値を計算する。

【 0 0 3 0 】

理論値計算部 1 1 による計算の手順は次のとおりである。すなわち、選択された制限酵素の切断箇所と切断位置のデータを制限酵素データ 2 1 から読み取り、この切断箇所の塩基配列パターンとマッチする部分を選択された増幅後塩基配列データ 2 2 b の塩基配列 (ORIGIN) 内で検索する。切断箇所が見つければ、始端からあるいは前回切断位置から今回切断位置までの塩基配列数を切断長理論値データ 2 3 に出力し、さらに検索を繰り返す。終端まで到達すれば、始端からあるいは前回切断位置から終端までの塩基配列数を切断長理論値データ 2 3 に出力する。

【 0 0 3 1 】

図 6 は、切断長理論値データ 2 3 のデータ構造およびデータ例を示す表図である。図 6 に示すように、切断長理論値データ 2 3 は、DNA 番号、数値区分、細菌名、遺伝子名、制限酵素番号切断長の列を有する。また切断長理論値データ 2 3 の各行は、DNA 毎 (DNA 番号により識別) かつ制限酵素毎 (制限酵素番号により識別) かつ切断長毎に存在する。ただし、切断片を識別するデータを列として加えて、同一 DNA の同一制限酵素による同一切断長を有する切断片を区別するようにしても良い。

【 0 0 3 2 】

これらの列のうち、DNA 番号と細菌名と遺伝子名の値は、理論値計算部 1 6 によって増幅後塩基配列データ 2 2 b から読み取られ出力されたものである。また、制限酵素番号の値は、同じく制限酵素データ 2 1 から読み取られ出力されたものである。また、切断長の値は理論値計算部 1 1 によって前記手順で算出されたものである。なお、数値区分は理論値か実測値かの区分を示すデータであり「1」は理論値であることを表している。

【 0 0 3 3 】

次に、切断長実測値データ 2 5 の詳細について説明する。切断長を実測するた

め、塩基配列の特定の対象となる試料は、予めプライマによって増幅されてから、制限酵素の適用によって切断される。このようにして作られた切断片の長さの測定は従来技術によって可能であり、例えば電気泳動によって得られる画像を数値化することにより行う。より具体的には、例えば、アガロースゲル電気泳動により切断片長分布を画像として得るようにする。

【 0 0 3 4 】

上記のような方法で実測された切断長は、実測値入力部 1 2 から入力される。図 7 は、実測値入力部 1 2 の利用者インタフェース例を示す画面図である。図 7 に示す画面では、利用者は、試料を識別するための DNA 番号を入力し、制限酵素データ 2 1 を基に表示される制限酵素のリストから 1 つを選択し、実測結果の切断長を入力することができる。また、複数の試料をグルーピングして付与したグループ名を入力することもできる。

【 0 0 3 5 】

図 8 は、切断長実測値データ 2 5 のデータ構造およびデータ例を示す表図である。切断長実測値データは、図 8 に示すように表形式のデータであり、DNA 番号、数値区分、グループ名、制限酵素番号、切断長の列を有する。また切断長実測値データ 2 5 の各行は、試料毎（DNA 番号により識別）かつ制限酵素毎（制限酵素番号により識別）かつ切断長毎に存在する。なお、数値区分は理論値か実測値かの区分を示すデータであり「2」は実測値であることを表している。

【 0 0 3 6 】

次に、比較部 1 3 が理論値と実測値との比較を行い両者の類縁性を分析する方法について説明する。本実施形態においては、比較部 1 3 は、非加重結合法（UPGMA）を用いることによって、解析結果データ 2 7 を生成する。

【 0 0 3 7 】

図 9 は、比較部 1 3 による塩基配列のクラスタリングの基となる切断片長のバンド分布を表す参考図である。図 9 において、A と B と C は、それぞれ DNA 番号によって識別される塩基配列の理論値または実測値の切断長分布である。このバンド分布は、電気泳動によって得られる画像パターンと同様のものであり、縦方向が切断長の次元である。また図 9 では、実測値データだけでなく理論値デー

タについても、得られるはずの仮想的な電気泳動パターンを示している。

【0038】

比較部13は、これらの実測値データおよび理論値データを読み込み、2つのDNA番号で与えられるデータ間の類似度を算出する。AとBの間の類似度S（A，B）の定義は、次式で与えられる。

【0039】

【数1】

$$S(A,B) = \frac{\text{AとBに共通するバンドの総数}}{\text{Aのバンド数} + \text{Bのバンド数}}$$

【0040】

例えば、図9に示すAとBでは、塩基配列Aには1，2，4，5，8の5本のバンドがあり、塩基配列Bには1，3，4，5，7，8の6本のバンドがある。従ってAとBに共通するバンドは、AおよびBそれぞれに1，4，5，8の4本であり、両方をたすと共通するバンドは合計8本である。従って、

【数2】

$$S(A,B) = \frac{4 \times 2}{6 + 5}$$

であり、類似度は8／11（約0．727）と算出される。

【0041】

ただし、実測値データは誤差を含むため、共通のバンドであるかどうかの判定にあたっては、比較部13は許容誤差データ29を用いることとする。つまり、設定された許容誤差範囲内であれば共通のバンドとして扱う。なお、設定される許容誤差が小さすぎると、本来マッチすべき切断長実測値と切断長理論値とがうまくマッチしないという問題が起こる。また逆に、設定される許容誤差が大きすぎると、切断長実測値が本来マッチすべきでない切断長理論値とマッチしてしまい、マッチングの精度が悪くなるという問題が起こる。従ってこれらの問題が起

こらないように適切な許容誤差を設定する必要がある、例えば、10%程度の許容誤差を設定するようにする。

【0042】

次に、比較部13は、算出した類似度を基に塩基配列のクラスタ分析を行う。図10は、比較部13が塩基配列間の類似度を基に平均距離法を用いてクラスタ分析を行う過程の例を示す表図であり、この図に示す例ではA、B、C、Dの4つの塩基配列を分析対象としている。図10(a)は、算出されたA、B、C、D相互間の類似度を示している。図10(a)の表の中で最も高い類似度を持っているのはAとC（類似度0.8）であるため、これらをまとめて(A+C)のクラスタとして次のステップに進む。

【0043】

図10(b)は、図10(a)のAおよびCを単一のクラスタ(A+C)で置き換えたときの類似度を示している。ここで、塩基配列Bとクラスタ(A+C)との類似度 $S(B, A+C)$ は、 $S(B, A)$ と $S(B, C)$ との平均により0.6であり、塩基配列Dとクラスタ(A+C)との類似度 $S(D, A+C)$ は、 $S(D, A)$ と $S(D, C)$ との平均により0.2となっている。そして、この表の中で最も高い類似度を持っているのは(A+C)とBであるため、これらをまとめて((A+C)+B)のクラスタとして次のステップに進む。

【0044】

図10(c)は、図10(b)の(A+C)およびBを単一のクラスタ((A+C)+B)で置き換えたときの類似度を示している。ここで、クラスタ((A+C)+B)と塩基配列Dとの類似度 $S(D, (A+C)+B)$ は、 $S(D, A+C)$ と $S(D, B)$ との平均により0.3となっている。

【0045】

図11は、上に説明した非加重結合法による解析結果データを、木構造を用いて図形として表した参考図である。図11において、木のノード部分に付記された数値は、そのノードに従属する終端ノード（塩基配列を表す）あるいは非終端ノード（クラスタを表す）間の類似度を表している。例えば、塩基配列Bとクラスタ(A+C)との類似度は0.6である。

【 0 0 4 6 】

表示部 1 5 は、解析結果データ 2 7 を読み込み、与えられた試料と類似度の高い既知遺伝子とを関連付けて図形的に表示する。図 1 2 は、表示部 1 5 によって樹形図として表示されている解析結果の例を示す画面図である。図 1 2 の例では、複数の試料 DNA (A P 1 , A P 2 , A P 3) と複数の既知遺伝子とのクラスタ分析の結果が示されている。

【 0 0 4 7 】

次に、本発明の第二実施形態について説明する。この第二実施形態は、増幅前の塩基配列を表すデータ内において、プライマによって増幅される塩基配列部分を自動的に認識して抽出する機能を有することを特徴とする。図 1 3 は、この第二実施形態による塩基配列解析システムの構成を示すブロック図である。図 1 3 において、符号 2 2 a はプライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データであり、3 1 は増幅前塩基配列データ 2 2 a を読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データ 2 2 b を生成する増幅配列認識部である。

【 0 0 4 8 】

図 1 4 は増幅配列認識部 3 1 への入力データとなる増幅前塩基配列データ 2 2 a の前半部分を示し、また、配列番号 2 は同データ 2 2 a の後半部分を示している (図 1 5 参照) 。図示するように、増幅前塩基配列データ 2 2 a は増幅後塩基配列データ 2 2 b と同様のテキスト形式のデータである。図 3 および図 1 5 は、同一の遺伝子 (ACCESSION = " M59070 " , KEYWORDS = " 16S ribosomal RNA . ") のものであるが、増幅後の塩基配列 (ORIGIN) は、増幅前の塩基配列 (ORIGIN) の 2 3 番目の文字から 1 0 3 1 番目の文字までの部分列であり、その長さは 1 0 0 9 となっている。

【 0 0 4 9 】

図 1 6 は、増幅配列認識部 3 1 による認識のためのプライマ配列の入力画面を示す画面図である。この画面では、上流プライマおよび下流プライマそれぞれについて、利用者がその配列と名前とミスマッチの許容限界を入力できるようにな

っている。図16の例では、上流プライマとして「41f」という名前で配列番号3の「gctcagattgaactcggcg」という配列が入力されており、下流プライマとして「1066r」という名前で配列番号4の「acatttcacacacgagctg」という配列が入力されている。そして、入力されたこれらの塩基配列パターンはプライマデータとして増幅配列認識部31によって利用される。

【0050】

増幅配列認識部31は、増幅前塩基配列データ22aを読み込み、その塩基配列（ORIGIN）を、始端から順に走査する。そして、上流プライマおよび下流プライマそれぞれについて、ミスマッチの許容限界として指定された数値範囲内でマッチする箇所を探し出す。そしてその箇所よりも上流および下流をそれぞれ切断し、残った塩基配列を増幅後塩基配列データ22bの塩基配列（ORIGIN）として出力する。

【0051】

上述の遺伝子の類縁性検索システムはコンピュータシステムを用いて実現されている。そして、上述した理論値計算部、実測値入力部、比較部、許容誤差設定部、表示部の各々の処理の過程は、コンピュータプログラムの形式でコンピュータ読み取り可能な記録媒体に記憶されており、このコンピュータプログラムをコンピュータが読み出して実行することによって、上記処理が行われる。ここでコンピュータ読み取り可能な記録媒体とは、フロッピーディスク、光磁気ディスク、CD-ROM、DVD-ROM、磁気ハードディスク、半導体メモリ等をいう。

【0052】

なお、この遺伝子の類縁性検索システムを実現するためのコンピュータシステムとしては、パーソナルコンピュータやワークステーションなどの汎用のコンピュータを用いることが可能である。

【0053】

また、遺伝子の類縁性検索システムにおいて扱うデータの形式は、上記第一および第二実施形態において記載したものに限定されず、他の形式あるいは表現の

データを用いても良い。また、上記実施形態で説明した表形式のデータの一部は非正規形であるが、正規化しても良い。例えば、図6に示す切断長理論値データにはDNA番号と細菌名と遺伝子名の列が含まれているが、この3者の関係を別表に保持するようにしても良い。

【0054】

次に、本発明による遺伝子の類縁性検索システムを用いて、細菌の分類を行った実例について説明する。

【0055】

生物の分類において最も元となる門 (division) は植物、動物、糸条菌などの真核細胞を有し有性生殖を行う真核生物と、単細胞である真正細菌 (細菌)、古細菌の3つの門に分類されている。初期の細菌分類体系では、細菌の細胞形態や生理学的性質、そして生化学的活性等の基準を基に分類体系が出来上がってきた。複雑で煩雑な操作が要求される菌の同定を容易に実施できるように、従来技術によるシステムにおいては、同定しようとする菌の複数の生理試験結果を自動的に読み取り、予め登録しデータベース化していた既知菌株の生理試験結果と比較して類縁性を検索し菌の種名を推定していた。このような生理学的試験に基づく同定システムは、医学・臨床分野や、化粧品、食品衛生、品質管理、環境衛生等の幅広い産業分野で利用されてきている。

【0056】

しかしながら、近年は、リボソームRNAを分子時計とする系統分類体系が構築され、細菌の16S rDNAの塩基配列の相違に基づく分類体系に従い分類体系が再構築されてきている。この理由は有性生殖を行う真核生物と異なり細胞分裂で増殖する真正細菌、古細菌では明確な種の概念が設定し難く、リボソームRNAを分子時計とする系統分類体系を基準とすることが定まったためである。旧来の生理試験結果に基づく同定法は、必ずしも現在の分類体系を正確に反映させることが困難であり、リボソームRNAを分子時計とする系統分類体系を正確に反映した本例のような簡易分類・同定手法が必要とされる。

【0057】

なお、細菌の名前は属 (genus) と種 (species) の二名式命名法で表記される

が、高順位の範疇として科 (family)、目 (class)、綱 (division)、群 (group) と真正細菌門 (division) の下に順次まとめられている。

【0058】

本例では、次のような実験を行った。

試料AP1～AP9として、様々な土壌から分離したアルキルフェノール分解菌（九州共立大学、名城大学）を用いる。また、試料MA1～MA4として、アルカリ耐性菌（山口大学）を用いる。そして、定法に従って染色体DNAを各菌株から抽出し、PCR反応を行い、増幅した16S rDNAを制限酵素で切断後各切断片の長さをデンストグラムで読み取り、実測値切断長として入力した。なお、試料MA1と試料MA11とは、同じ菌株の同じDNAを用いてこれらの操作を別途行ったものである。

【0059】

図17は、これらの試料から得られた実測値同士の類縁性を、本発明による遺伝子の類縁性検索システムでの解析で検定した結果を示す樹形図である。

【0060】

また、図18～図27は、それぞれ、異なった試料の菌株から得られた実測値と遺伝子配列データを基に得られた理論値との類縁性の検定を行った結果を示す樹形図である。なお、図18は試料AP1について、図19は試料AP3について、図20は試料AP6について、図21は試料AP2について、図22は試料AP5について、図23は試料MA1について、図24は試料MA11について、図25は試料MA2について、図26は試料MA3について、図27は試料MA4について、それぞれ上記検定を行った結果を示す。

【0061】

上記検定を行ったとき、切断長理論値データ（23）としては、357属、1233種、1503種類のDNA配列データから作成した理論値データが登録されていた。この理論値データは、すべて公開されているDNA配列データを基に本発明による遺伝子の類縁性検索システムによって算出されたものであり、細菌の種類のカテゴリ・同定の基礎とするのに十分なこれほど多種のデータを安価かつ短時間で作成することができたのは、本発明を用いたことによる大きな効果である。

。これに対して、例えば、従来技術による生理学的試験に基づくある同定キット（市販品）の場合には、数十年にもおよぶ長い年月をかけてデータを整備することにより、ようやく1210種、200属に満たない細菌を同定できるようになっているに過ぎない。また、他のある同定キット（市販品）の場合にも、同様に約700菌種を同定できるようになっているに過ぎない。これらの従来技術の方法に基づくシステムと比べて、本発明によるシステムは、今後新たに追加されるDNA配列データも含めて、公開データを基に極めて安価かつ短時間でデータを増やすことができるという利点を持っている。

【0062】

前記検定結果によると、試料MA1および試料MA11は、*Aeromonas hydrophila* (Proteobacteria α subdivision, *Aeromonas* group) に最も近く、同じProteobacteria α subdivisionに属するが少し離れたEnterobacteriaceae（腸内細菌科）に属するSerratia属、Shewanella属、またはPlesiomonas属が近いという結果が出ている。

【0063】

また、試料MA2は、*Sporolactobacillus*属、または*Bacillus*属（いずれもLowGC gram positive bacteria、*Bacillus*/*Clostridium*、*Bacillaceae* (*Bacillus* 科)) に等しく近いという結果が得られている。

【0064】

また一方で、試料MA3は、*Bacillus cereus*と*Bacillus thuringiensis* (LowGC gram positive bacteria、*Bacillus*/*Clostridium*、*Bacillaceae*) に最も近いという結果が得られている。なお、*Bacillus cereus*と*Bacillus thuringiensis*は近縁であり同一種とする意見が多い。

【0065】

また、試料MA4は、*Oerskovia*属もしくは*Cellulomonas*属(共にActinobacteria、Actinobacteridae、Actinomycetales、Micrococcineae、Cellulomonadaceae) に最も近く、Actinomycetes属(Actinobacteria、Actinobacteridae、Actinomycetales、Actinomycetaceae)、*Streptomyces*属 (Actinobacteria、Actinobacteridae、Actinomycetales、Streptomyces、Streptomycetaceae)

) に次に近い。

【 0 0 6 6 】

試料 A P 1 と A P 3 は、*Pseudomonas putida*, *Pseudomonas fulva*, *Pseudomonas straminea*, *Pseudomonas alcaligenes*, *Flavimonas oryzihabitans* (いずれも *Proteobacteria* γ subdivision, *Pseudomonadaceae*) に近い。

【 0 0 6 7 】

また、試料 A P 6 は、*Xanthomonas* 属細菌 (*Proteobacteria* γ subdivision, *Lysobacterales*, *Xanthomonas* group) に近く、試料 A P 2 と A P 5 は、*Phyllobacterium* 属、*Rhizobium* 属、*Agrobacterium* 属 (*Proteobacteria* α subdivision, *Rhizobiceae* group のそれぞれ *Phyllobacteriaceae* か *Rhizobiaceae*) と推定された。

【 0 0 6 8 】

図 2 8 は、分離した脱窒菌 1 3 4 株について本方法で種名を推定し代表的な菌株の 1 6 S r D N A の部分塩基配列を決定し、公開用塩基配列データベースで相同性検索した結果最も近い 1 6 S r D N A の塩基配列構造を有する菌の種名を検索した結果を示している。

【 0 0 6 9 】

この図 2 8 では、分類されたグループ (I から XIII) 、そのグループに含まれる菌数、推定される種名 (RFLP) 、1 6 S r D N A の塩基配列を実際に決定し、最も近い塩基配列を有する菌の名前および相違度を示している。この場合、類似度が高い塩基配列部分の塩基配列のうち、いくつ一致したかを百分率表示で示している。複数の数値が示されているのは、そのグループに属する別の細菌の塩基配列を決定し相同性検索を別に行った結果である。

【 0 0 7 0 】

【発明の効果】

以上説明したように、この発明によると、既知の遺伝子の塩基配列と制限酵素の切断パターンとを基に予め切断長理論値データを作成しておき、この切断長理論値データと入力される切断長実測値データとを非加重結合法などを用いて分析するため、既知遺伝子と試料との類似度を高速かつ効率的に算出することが可能

となる。

【 0 0 7 1 】

また、この発明によると、多種の既知遺伝子および試料との類似度を基に、類似度の高いもの同士を関連付けて、樹形図などを用いて図形的に表示するため、解析結果を利用者に理解しやすい形で提供することが可能となる。

【 0 0 7 2 】

また、この発明によると、増幅配列認識部が増幅前塩基配列データを基に、増幅部分を自動的に認識して増幅後塩基配列データを生成するため、多くの手間をかけることなく増幅後塩基配列データを準備することができ、塩基配列の解析をより一層効率化することが可能となる。

【 0 0 7 3 】

また、この発明によると、パーソナルコンピュータやワークステーションなどの汎用のコンピュータを用いて遺伝子の類縁性検索システムを実現することにより、専用のシーケンサ等に比べて極めて安価なシステムを提供することが可能となる。

【 0 0 7 4 】

【配列表】

< 1 1 0 > Director General of Kyushu National Agricultural Experiment Station; Katsuji Watanabe

< 1 2 0 > Methods and systems for analyzing genetic relationship

< 1 6 0 > 4

< 2 1 0 > 1

< 2 1 1 > 1 0 0 9

< 2 1 2 > DNA

< 2 1 3 > Rhodospirillum salexigens

< 2 2 0 >

< 2 2 1 > r RNA

< 2 2 2 > (50), (51), (63), (463), (468), (574), (687), (806), (853), (854), (888), (889), (966)

< 2 2 3 > u n k n o w n

< 4 0 0 >

1 gctcagaacg aacgctggcg gcaggcctaa cacatgcaag tcgagcgcan nccttcgggg
 61 gtnagcggcg gacgggtgag taacgcgtgg gaacctgctc agggctctgg gataactgct
 121 ggaaacggca gctaataccg gatacgccgt attgggaaag aaattcggcc ttggatgggc
 181 ccgcgttggg ttagctagat ggtggggtaa cggcctacca tggcgacgat ccatagctgg
 241 tttgagagga tgatcagcca cactgggact gagacacggc ccagactcct acgggaggca
 301 gcagtgggga atcttagaca atgggggcaa ccctgatcta gccatgccgc gtgagtgatg
 361 aaggccttag ggttgtaaag ctctttcagc agggaagata atgactgtac ctgcagaaga
 421 agctccggct aactccgtgc cagcagccgc ggtaatacgg agngggcnag cgttgttcgg
 481 aattactggg cgtaaagcgc gcgtaggcgg atcggtcagt tgggggtgaa agcccggggc
 541 tcaacctcgg aactgccctc aaaactaccg atcnagagtt cgggagaggt aagcgggaatt
 601 cccagtgtag aggtgaaatt cgtagatatt gggaagaaca ccagtggcga aggcggctta
 661 ctggaccgat actgacgctg aggtgcnaaa gcgtggggag caaacaggat tagataccct
 721 ggtagtccac gccgtaaacg atgggtgcta gatgtcgggg ctcttagagt ttcggtatcg
 781 cagctaacgc attaagcacc ccgccngggg agtacggccg caagggtaaa actcaaagga
 841 attgacgggg gcnngcacaa gcggtggagc atgtggttta attcgaanna acgcgcagaa
 901 ccttaccagc tcttgacatc ccgggacgac ttccagagat ggattttttc acttcggtga
 961 cccgngaca ggtgctgcat ggctgtcgtc agctcgtgtc gtgagatgt

< 2 1 0 > 2

< 2 1 1 > 1 4 9 0

< 2 1 2 > D N A

< 2 1 3 > *Rhodospirillum salexigens*

< 2 2 0 >

< 2 2 1 > r R N A

< 2 2 2 > (1), (72), (73), (85), (485), (490), (596), (709), (828), (875), (876), (910), (911), (988), (1045), (1332), (1333), (1408), (1437)-(1442), (1451)-(1475)

< 2 2 3 > u n k n o w n

< 4 0 0 >

1 ncaacatgag agtttgatcc tggctcagaa cgaacgctgg cggcaggcct aacacatgca
 61 agtcgagcgc anncccttcgg gggtnagcgg cggacgggtg agtaacgcgt gggaacctgc
 121 tcagggctct gggataactg ctggaaacgg cagctaatac cggatacgcc gtattgggaa
 181 agaaattcgg ccttgatgg gcccgcttg gattagctag atggtggggt aacggcctac
 241 catggcgacg atccatagct ggtttgagag gatgatcagc cacactggga ctgagacacg
 301 gcccagactc ctacgggagg cagcagtggg gaatcttaga caatgggggc aaccctgatc
 361 tagccatgcc gcgtgagtga tgaaggcctt agggttgtaa agctctttca gcagggaaga
 421 taatgactgt acctgcagaa gaagctccgg ctaactccgt gccagcagcc gcggtaatc
 481 ggagngggcn agcgttggtc ggaattactg ggcgtaaagc gcgcgtaggc ggatcgggtc
 541 gttgggggtg aaagcccggg gctcaacctc ggaactgccc tcaaaactac cgatcnagag
 601 ttcgggagag gtaagcggaa ttcccagtgt agagggtgaaa ttcgtagata ttgggaagaa
 661 caccagtggc gaaggcggct tactggaccg atactgacgc tgagggtgna aagcgtgggg
 721 agcaaacagg attagatacc ctggtagtcc acgccgtaaa cgatgggtgc tagatgtcgg
 781 ggctcttaga gtttcggtat cgcagctaac gcattaagca ccccgccngg ggagtacggc
 841 cgcaagggtta aaactcaaag gaattgacgg gggcnngcac aagcgggtga gcatgtggtt
 901 taattcgaan naacgcgcag aaccttacca gctcttgaca tcccgggacg acttcagag
 961 atggattttt tcaattcggg gacccggnga cagggtgctgc atggctgtcg tcagctcgtg
 1021 tcgtgagatg ttgggttaag tcccncaacg agcgcaaccc tcgcccttag ttgccagcat
 1081 ttggttgggg actctaaggg aactgccggt gataagccgg aggaagggtg ggatgacgtc
 1141 aagtcctcat ggcccttatg ggctgggcta cacacgtgct acaatggcgg tgacagaggg
 1201 cagcgagcct gcgagggtga gcgaatctct aaaagccgtc tcagttcgga ttgttctctg
 1261 caactcgaga gcatgaagggt ggaatcgcta gtaatcgcg atcagcatgc cgcggtgaat
 1321 acgttcccgg gnnttgatca caccgccgt cacaccatgg gagttggtt gaccgaaga
 1381 cggtagacta acccgaaagg ggggcagncg gccacggtca ggtagcgac tggggtnnnn
 1441 nngtaacaag nnnnnnnnnn nnnnnnnnnn nnnngatca cctcctttct

< 2 1 0 > 3

< 2 1 1 > 1 9

< 2 1 2 > DNA

< 2 1 3 > A r t i f i c i a l S e q u e n c e

< 2 2 0 >

< 2 2 3 > A r t i f i c i a l S e q u e n c e

< 4 0 0 >

1 g c t c a g a t t g a a c t c g g c g

< 2 1 0 > 4

< 2 1 1 > 1 6

< 2 1 2 > DNA

< 2 1 3 > A r t i f i c i a l S e q u e n c e

< 2 2 0 >

< 2 2 3 > A r t i f i c i a l S e q u e n c e

< 4 0 0 >

1 a c a t t t c a c a a c a c g a g c t g

【図面の簡単な説明】

【図 1】 この発明の第一実施形態による塩基配列解析システムの構成を示すブロック図である。

【図 2】 同実施形態による増幅後塩基配列データ（2 2 b）の前半部分を示す図である。

【図 3】 同実施形態による増幅後塩基配列データ（2 2 b）の後半部分を示す図である。

【図 4】 同実施形態による制限酵素データ（2 1）のデータ構造およびデータ例を示す表図である。

【図 5】 同実施形態による理論値計算部（1 1）に対して切断長理論値計算を指示する利用者インタフェース例を示す画面図である。

【図 6】 同実施形態による切断長理論値データ（2 3）のデータ構造およびデータ例を示す表図である。

【図 7】 同実施形態による実測値入力部（12）の利用者インタフェース例を示す画面図である。

【図 8】 同実施形態による切断長実測値データ（25）のデータ構造およびデータ例を示す表図である。

【図 9】 同実施形態による塩基配列のクラスタリングの基となる切断片のバンド分布を表す参考図である。

【図 10】 同実施形態により塩基配列間の類似度を基に平均距離法を用いてクラスタ分析を行う過程を示す表図である。

【図 11】 同実施形態によるクラスタ分析の結果を図形化して示した参考図である。

【図 12】 同実施形態による表示部（15）が解析結果を樹形図として表示する例を示す画面図である。

【図 13】 この発明の第二実施形態による塩基配列解析システムの構成を示すブロック図である。

【図 14】 同実施形態による塩基配列解析システムの入力データとなる増幅前塩基配列データ（22a）の前半部分を示す図である。

【図 15】 同実施形態による塩基配列解析システムの入力データとなる増幅前塩基配列データ（22a）の後半部分を示す図である。

【図 16】 同実施形態による増幅配列認識部（31）による認識のためのプライマ配列の入力画面を示す画面図である。

【図 17】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図 18】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図 19】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図 20】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図 21】 本発明による塩基配列解析システムを用いて、細菌の種の分類

- ・ 同定を行った例の結果を示す樹形図である。

【図 2 2】 本発明による塩基配列解析システムを用いて、細菌の種の分類

- ・ 同定を行った例の結果を示す樹形図である。

【図 2 3】 本発明による塩基配列解析システムを用いて、細菌の種の分類

- ・ 同定を行った例の結果を示す樹形図である。

【図 2 4】 本発明による塩基配列解析システムを用いて、細菌の種の分類

- ・ 同定を行った例の結果を示す樹形図である。

【図 2 5】 本発明による塩基配列解析システムを用いて、細菌の種の分類

- ・ 同定を行った例の結果を示す樹形図である。

【図 2 6】 本発明による塩基配列解析システムを用いて、細菌の種の分類

- ・ 同定を行った例の結果を示す樹形図である。

【図 2 7】 本発明による塩基配列解析システムを用いて、細菌の種の分類

- ・ 同定を行った例の結果を示す樹形図である。

【図 2 8】 分離した脱窒菌について本方法で種名を推定し代表的な菌株の 1 6 S r D N A の部分塩基配列を決定し、公開用塩基配列データベースで相同性検索した結果最も近い 1 6 S r D N A の塩基配列構造を有する菌の種名を検索した結果を示している表図である。

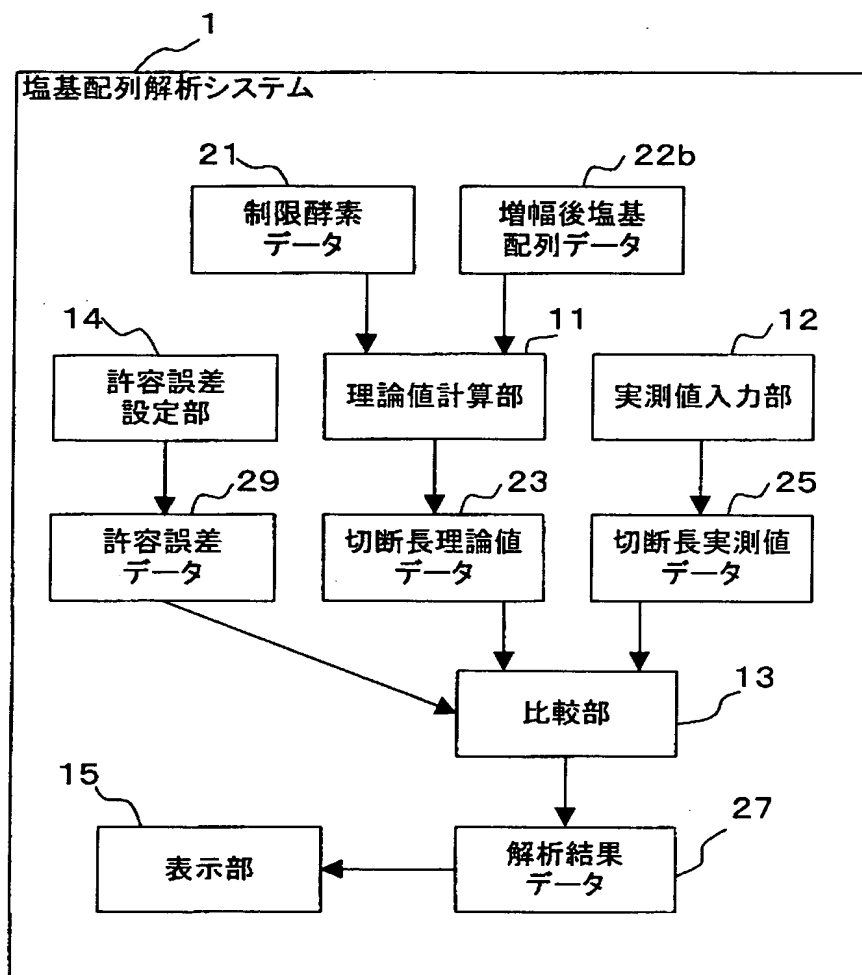
【符号の説明】

- 1 塩基配列解析システム
 - 1 1 理論値計算部
 - 1 2 実測値入力部
 - 1 3 比較部
 - 1 4 許容誤差設定部
 - 1 5 表示部
- 2 1 制限酵素データ
 - 2 2 a 増幅前塩基配列データ
 - 2 2 b 増幅後塩基配列データ
- 2 3 切断長理論値データ
- 2 5 切断長実測値データ

- 2 7 解析結果データ
- 2 9 許容誤差データ
- 3 1 増幅配列認識部

【書類名】 図面

【図 1】



【図 2】

LOCUS RSP16SRRZ 1490 bp rRNA BCT 04-DEC-1995
 DEFINITION Rhodospirillum salexigens 16S ribosomal RNA.
ACCESSION M59070
 NID g175871
 VERSION M59070.1 GI:175871
KEYWORDS 16S ribosomal RNA.
 SOURCE Rhodospirillum salexigens rRNA.
ORGANISM Rhodothalassium salexigens
Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
Rhodothalassium
 REFERENCE 1 (bases 1 to 1490)
 AUTHORS Woese, C. R.
 TITLE A phylogenetic analysis of the some purple bacteria
 JOURNAL Unpublished (1991)
 FEATURES Location/Qualifiers
 source 1..1490
 /organism="Rhodothalassium salexigens"
 /db_xref="taxon:1086"
 /tissue_lib="DSM 2132"
 rRNA 1..1490
 /gene="16S rRNA"
 /product="16S ribosomal RNA"
 gene 1..1490
 /gene="16S rRNA"
 BASE COUNT 342 a 343 c 472 g 284 t 49 others

【図 3】

ORIGIN

```

1 gctcagaacg aacgctggcg gcaggcctaa cacatgcaag tcgagcgcan nccttcgggg
61 gtnagcggcg gacgggtgag taacgcgtgg gaacctgctc agggctctgg gataactgct
121 ggaaacggca gctaataccg gatacgccgt attgggaaag aaattcggcc ttggatgggc
181 ccgcgttggg ttagctagat ggtggggtaa cggcctacca tggcgacgat ccatagctgg
241 ttgagagga tgatcagcca cactgggact gagacacggc ccagactcct acgggaggca
301 gcagtgggga atcttagaca atgggggcaa cctgatcta gccatgccgc gtgagtgatg
361 aaggccttag ggttgtaaag ctcttcagc agggaagata atgactgtac ctgcagaaga
421 agtccggct aactccgtgc cagcagccgc ggtaatacgg agngggcnag cgttgttcgg
481 aattactggg cgtaaagcgc gcgtaggcgg atcgggtcagt tgggggtgaa agccccgggc
541 tcaacctcgg aactgccctc aaaactaccg atcnagagtt cgggagaggt aagcgggaatt
601 cccagtgtag aggtgaaatt cgtagatatt gggaagaaca ccagtggcga aggcggctta
661 ctggaccgat actgacgctg aggtgcnaaa gcgtggggag caaacaggat tagataccct
721 ggtagtccac gccgtaaacg atgggtgcta gatgtcgggg ctcttagagt ttcggtatcg
781 cagctaacgc attaagcacc ccgccngggg agtaocggcg caaggttaaa actcaaagga
841 attgacgggg gcnngcacaa gcggtggagc atgtggttta attogaanna acgcgcagaa
901 ccttaccagc tcttgacatc ccgggacgac ttccagagat ggattttttc acttcggtga
961 cccggngaca ggtgctgcat ggctgtcgtc agctcgtgtc gtgagatgt

```

//

【図 4】

制限酵素データ

制限酵素番号	制限酵素名	切断箇所	切断位置
1	AluI	agct	2
2	HaeIII	ggcc	2
3	RsaI	gtac	2
4	ScrFI	ccngg	2
5	HhaI	gcgc	2
6	BamHI	ggatcc	1
7	EcoRI	gaatt	1
8	HindIII	aagctt	1
9	PstI	ctgcag	5
10	PvuII	cagctg	3
11	Sall	gagctc	5
12	SmaI	cccggg	3
13	XbaI	tctaga	1

【図 5】

TKS030

データベース登録(理論値)

塩基配列ファイル	制限酵素
AB021410.TXT	Acl
AF064459.TXT	HacI
AJ002302.TXT	Haa
L01165.TXT	RsaI
U96927.TXT	abd
X95918.TXT	
Y10657.TXT	
Z69293.TXT	

→

切断長計算 戻る

【図 6】

切断長理論値データ

DNA番号	数値区分	グループ名	細菌名	遺伝子名	制限酵素番号	切断長
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	2	204
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	5	509
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	5	542
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	3	194
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	3	457
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	3	400
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	2	156
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	2	90
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	2	317
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	2	169
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	1	550
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	1	180
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	1	79
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	1	87
AB000278	1		Vibrio iliopiscarius	16S rRNA; 16S ribosomal RNA	2	105
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	156
AB038030	1		Vibrio splendidus	16S ribosomal RNA	5	509
AB038030	1		Vibrio splendidus	16S ribosomal RNA	5	565
AB038030	1		Vibrio splendidus	16S ribosomal RNA	4	1074
AB038030	1		Vibrio splendidus	16S ribosomal RNA	3	194
AB038030	1		Vibrio splendidus	16S ribosomal RNA	3	234
AB038030	1		Vibrio splendidus	16S ribosomal RNA	3	223
AB038030	1		Vibrio splendidus	16S ribosomal RNA	3	423
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	90
AB038030	1		Vibrio splendidus	16S ribosomal RNA	1	175
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	204
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	317
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	105
AB038030	1		Vibrio splendidus	16S ribosomal RNA	1	236
AB038030	1		Vibrio splendidus	16S ribosomal RNA	1	572

【図 7】

TKS040

データベース登録(実測値)

分子量ファイル

Deni01(Ha).txt
DNA.xls

グループ名

AP

DNA番号

AP1

制限酵素

AluI

切断長

100	200	300					

登録

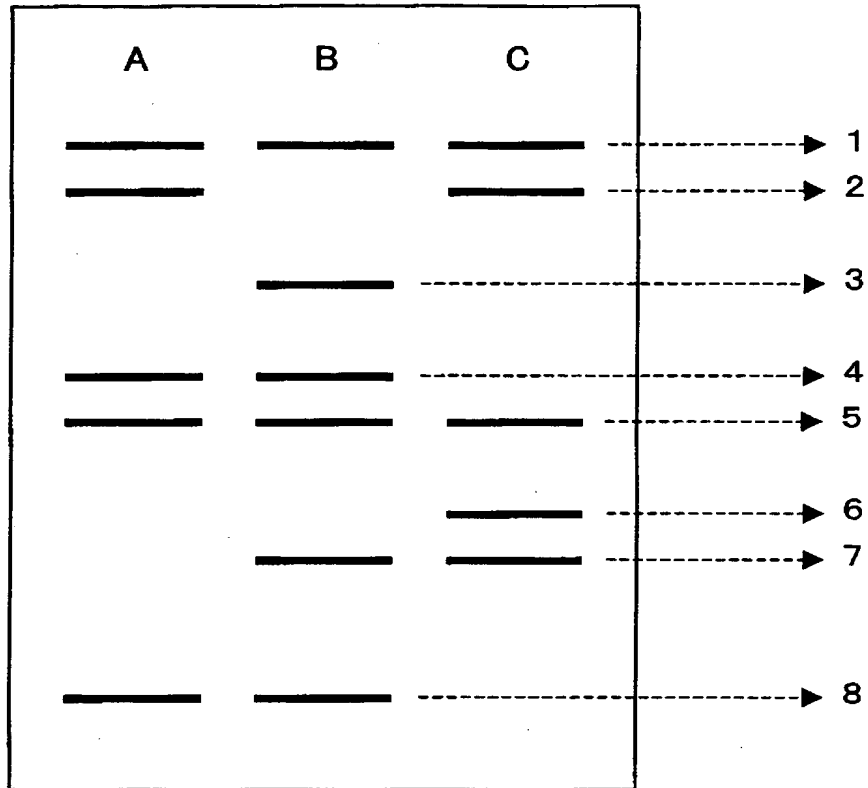
戻る

【図 8】

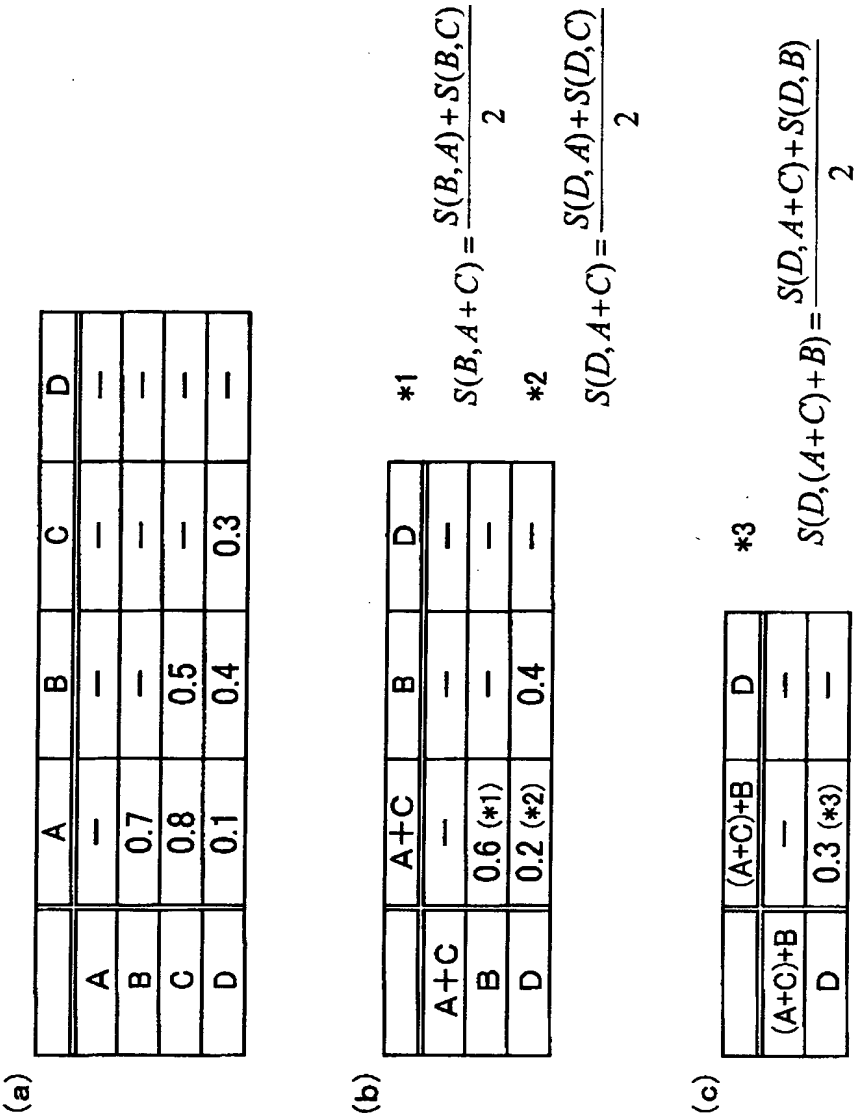
切断長実測値データ

DNA番号	数値区分	グループ名	細菌名	遺伝子名	制限酵素番号	切断長
AP1	2	AP			1	100
AP1	2	AP			1	200
AP1	2	AP			1	300

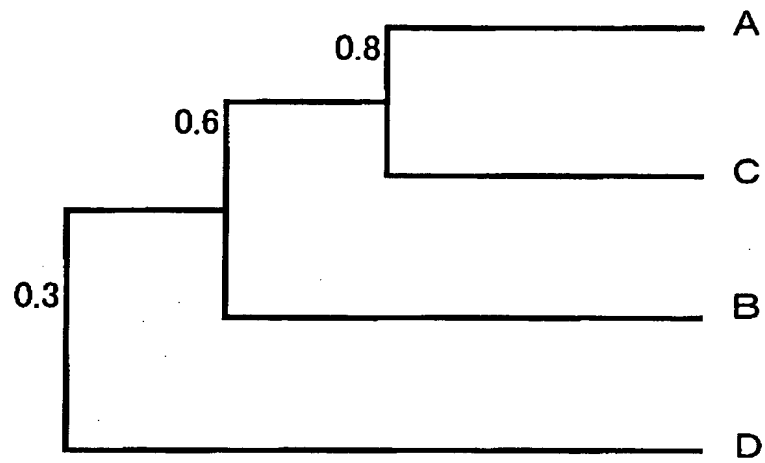
【図 9】



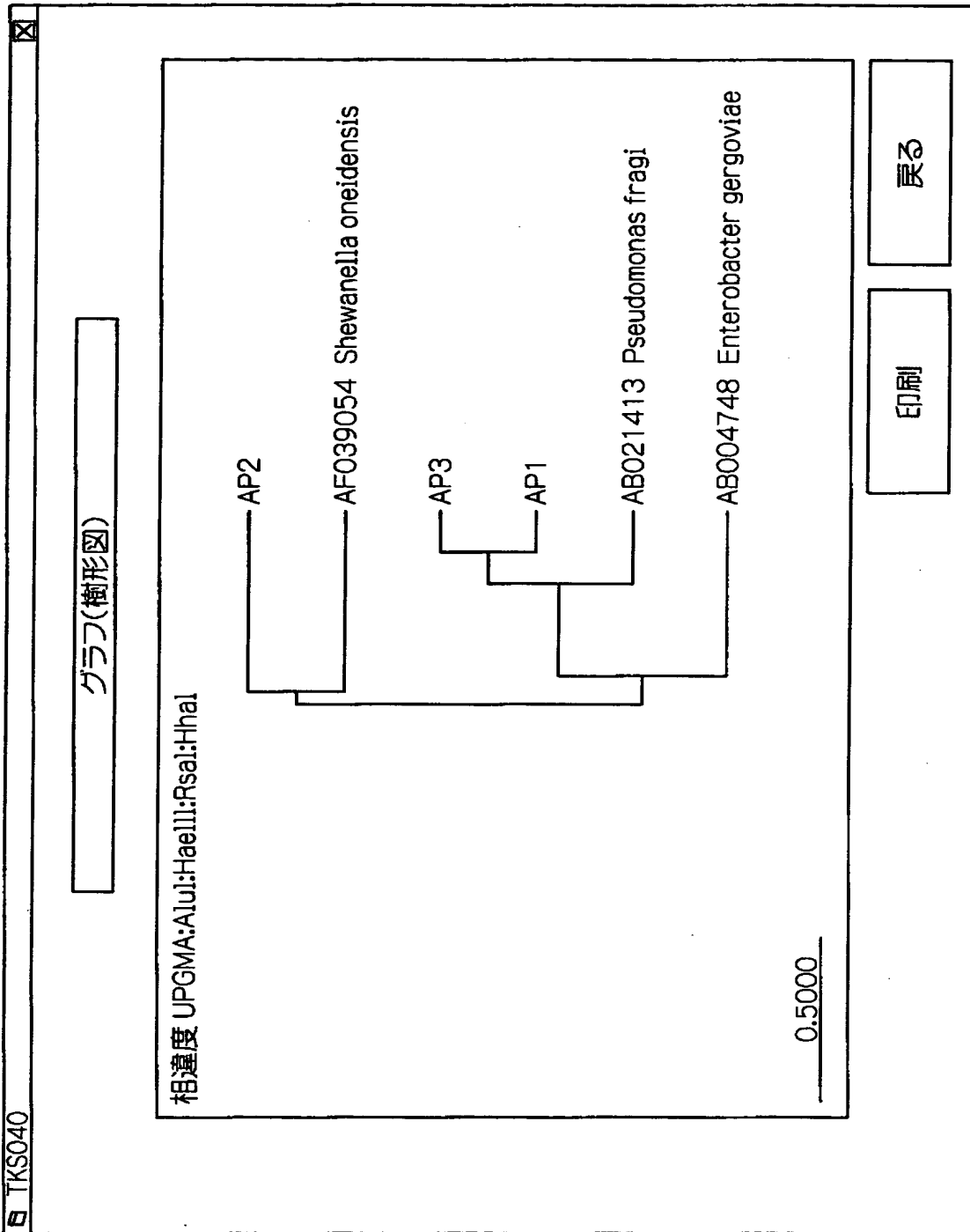
【図 1 0】



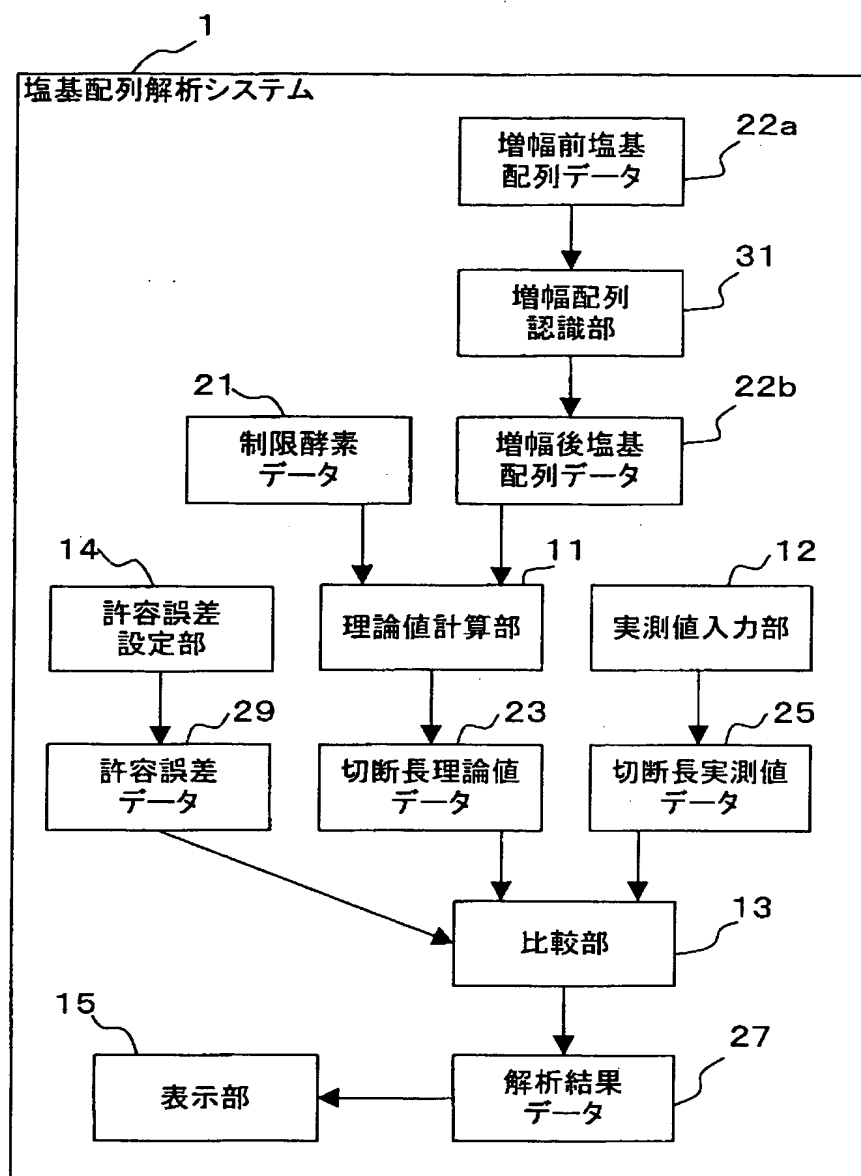
【図 1 1】



【図 12】



【図13】



【図 1 4】

LOCUS RSP16SRRZ 1490 bp rRNA BCT 04-DEC-1995
 DEFINITION Rhodospirillum salexigens 16S ribosomal RNA.
 ACCESSION M59070
 NID g175871
 VERSION M59070.1 GI:175871
 KEYWORDS 16S ribosomal RNA.
 SOURCE Rhodospirillum salexigens rRNA.
 ORGANISM Rhodothalassium salexigens
 Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 Rhodothalassium.
 REFERENCE 1 (bases 1 to 1490)
 AUTHORS Woese, C. R.
 TITLE A phylogenetic analysis of the some purple bacteria
 JOURNAL Unpublished (1991)
 FEATURES Location/Qualifiers
 source 1..1490
 /organism="Rhodothalassium salexigens"
 /db_xref="taxon:1086"
 /tissue_lib="DSM 2132"
 rRNA 1..1490
 /gene="16S rRNA"
 /product="16S ribosomal RNA"
 gene 1..1490
 /gene="16S rRNA"
 BASE COUNT 342 a 343 c 472 g 284 t 49 others

【図 1 5】

ORIGIN

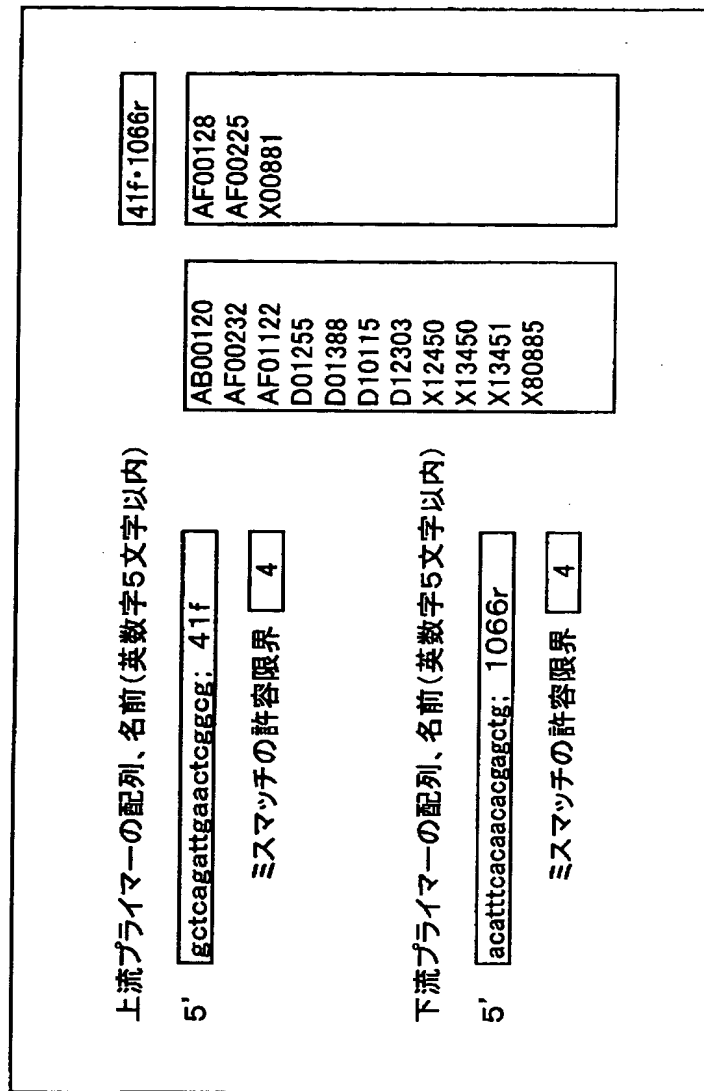
```

1 ncaacatgag agtttgatcc tggctcagaa cgaacgctgg cggcaggcct aacacatgca
61 agtcgagcgc annccitcgg gggtnagcgg cggacgggtg agtaacgcgt gggaacctgc
121 tcagggtctt gggataactg ctggaaacgg cagctaatac cggatacgcc gtattgggaa
181 agaaattcgg ccttggatgg gcccgcttg gattagctag atggtgggtt aacggcctac
241 catggcgacg atccatagct ggtttgagag gatgatcagc cacactggga ctgagacacg
301 gccagactc ctacgggagg cagcagtggg gaatcttaga caatgggggc aaccctgac
361 tagccatgcc gcgtgagtga tgaaggcctt agggttgtaa agctctttca gcagggaaga
421 taatgactgt acctgcagaa gaagctccgg ctaactccgt gccagcagcc gcggtaatac
481 ggagngggcn agcgttgttc ggaattactg ggcgtaaagc gcgcgtaggc ggatcggica
541 gttgggggtg aaagcccggg gctcaacctc ggaactgccc tcaaaactac cgatcnagag
601 ttcgggagag gtaagcggaa ttcccagtgt agaggtagaa ttcgtagata ttgggaagaa
661 caccagtggc gaaggcggct tactggaccg atactgacgc tgaggtagcna aagcgtgggg
721 agcaaacagg attagatacc ctggtagtcc acgccgtaaa cgatgggtgc tagatgtcgg
781 ggctcttaga gtttcggtat cgcagctaac gcattaaagc ccccgccngg ggagtacggc
841 cgcaaggtta aaactcaaag gaattgacgg gggcnngcac aagcggtigga gcatgtggtt
901 taattcgaan naacgcgcag aaccttacca gctcttgaca tcccgggacg acttccagag
961 atggattttt tcacttcggt gaccggnga cagggtgctg atggctgtcg tcagctcgtg
1021 tcgtgagatg ttgggttaag tccncaacg agcgcaacc tcgcccttag ttgccagcat
1081 ttggttgggg actctaaggg aactgccggt gataagccgg aggaaggtag ggatgacgtc
1141 aagtcctcat ggcccttatg ggctgggcta cacacgtgct acaatggcgg tgacagaggg
1201 cagcgagcct gcgagggtga gcgaatctct aaaagccgtc tcagttcgga ttgttctctg
1261 caactcgaga gcatgaagg ggaatcgcta gtaatcgcg atcagcatgc cgcggtgaat
1321 acgttcccgg gnnttgtaga caccgccgt cacaccatgg gagttggttt gaccogaaga
1381 cggtagagta acccgaaagg ggggcagnog gccacgtca ggtcagcgac tggggtnnnn
1441 nngtaacaag nnnnnnnnnn nnnnnnnnnn nnnngatca cctcctttct

```

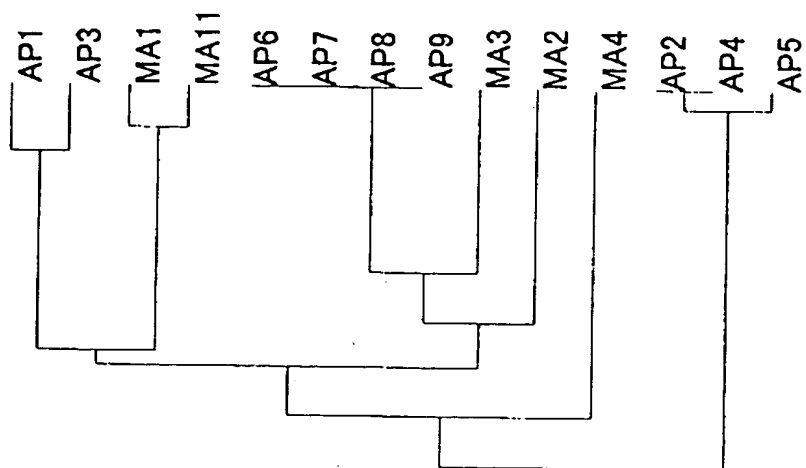
//

【図 16】



【図 1 7】

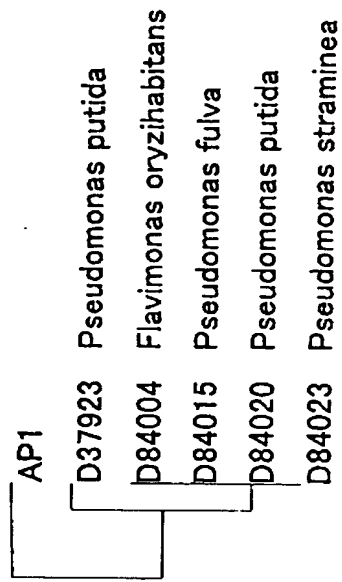
相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



0.5000

【図 1 8】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI

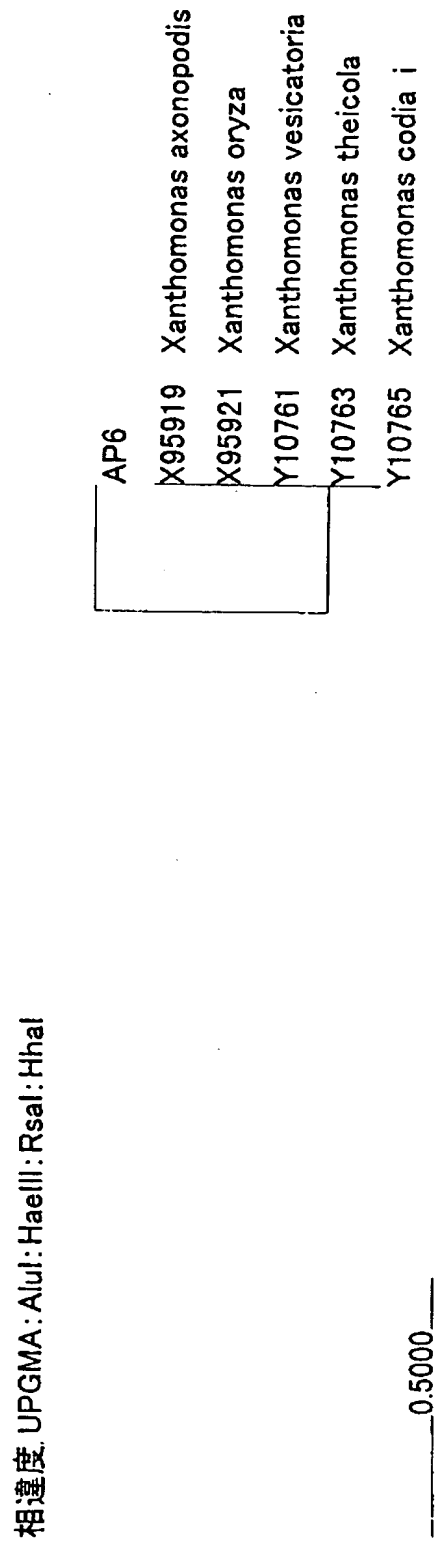


0.5000

【図 1 9】

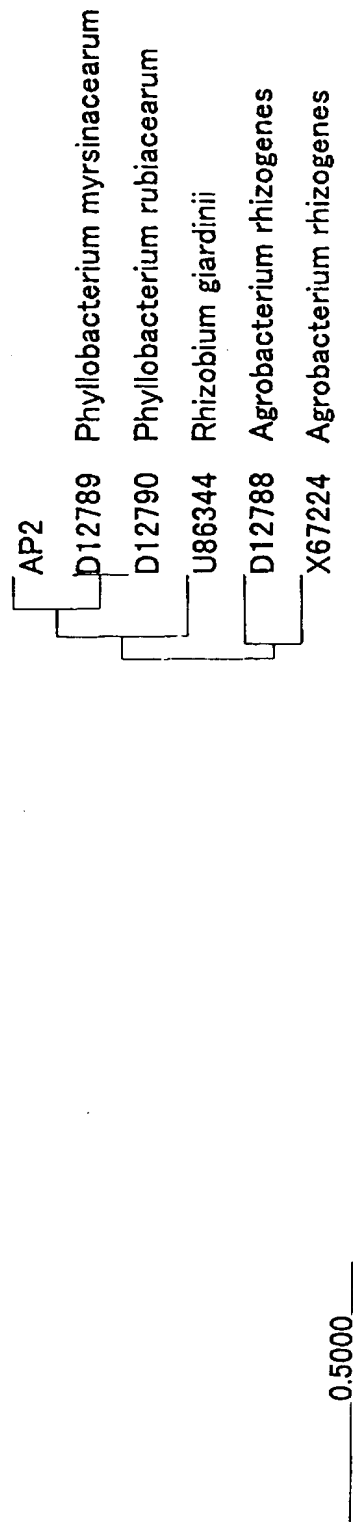


【図 2 0】

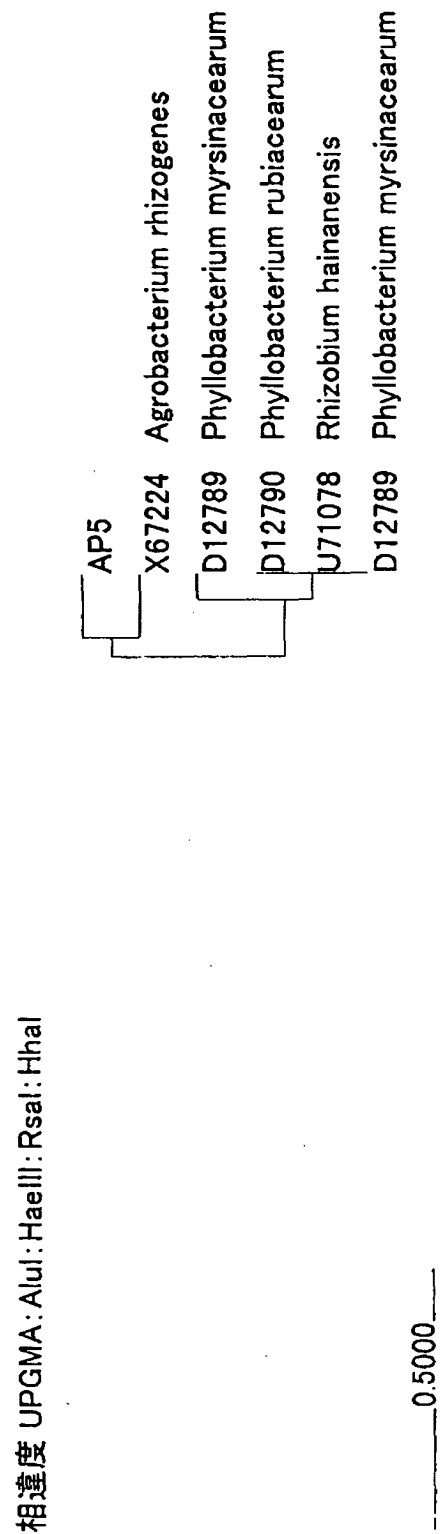


【図 2 1】

相違度 UPGMA: Alul: HaeIII: RsaI: HhaI

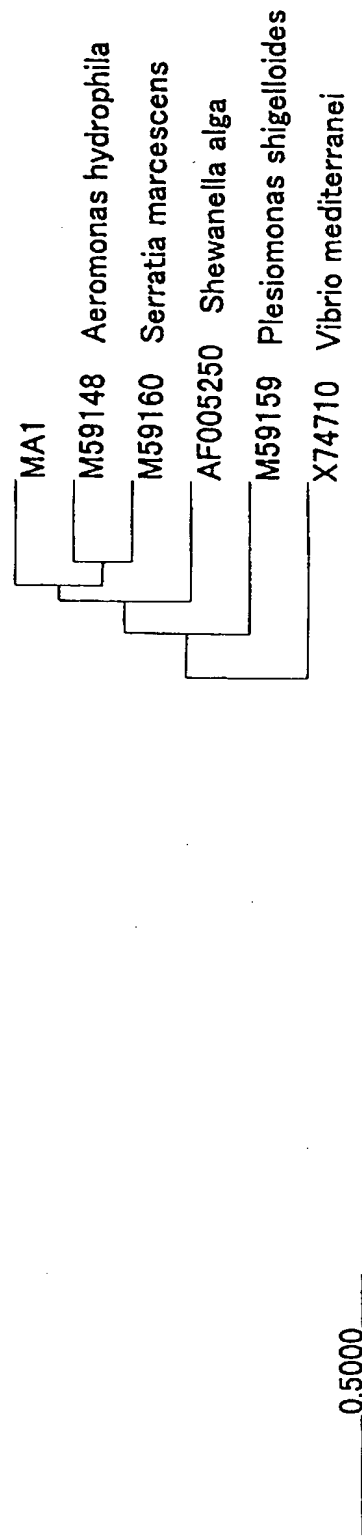


【図 2 2】



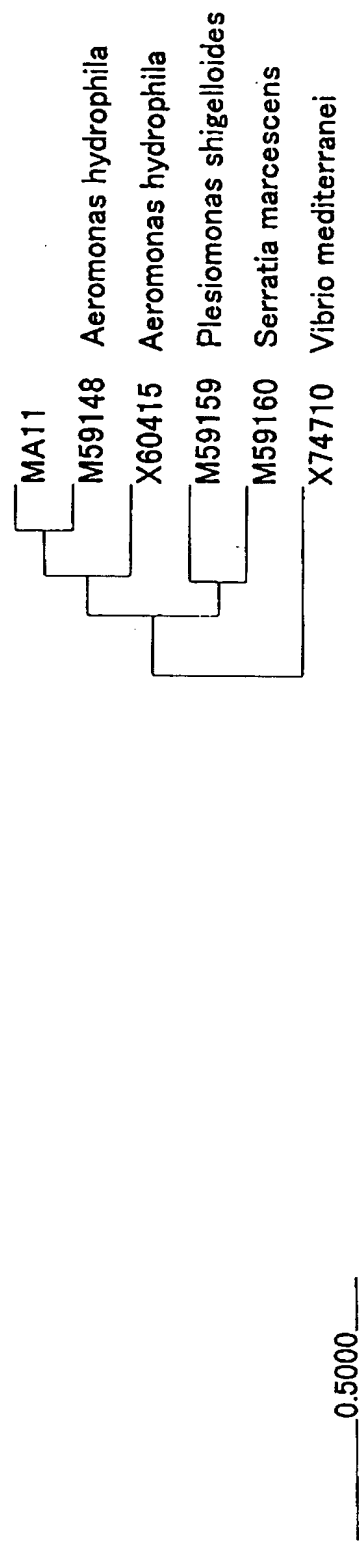
【図 2 3】

相違度 UPGMA: Alul: HaeIII: RsaI: HhaI

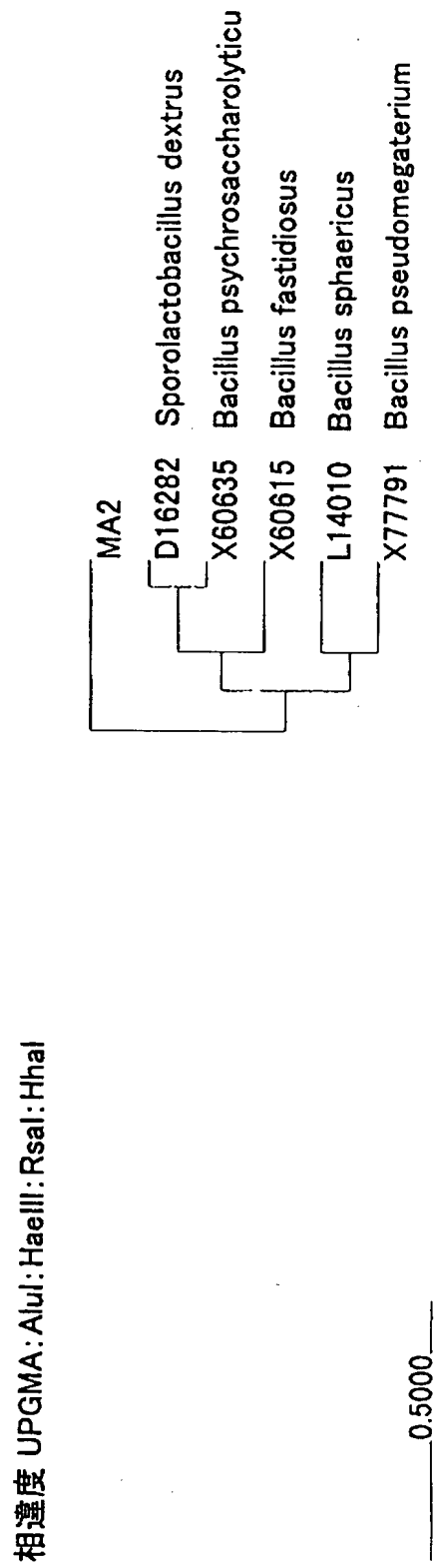


【図 2 4】

相違度 UPGMA: Alul: HaeIII: RsaI: HhaI



【図 2 5】

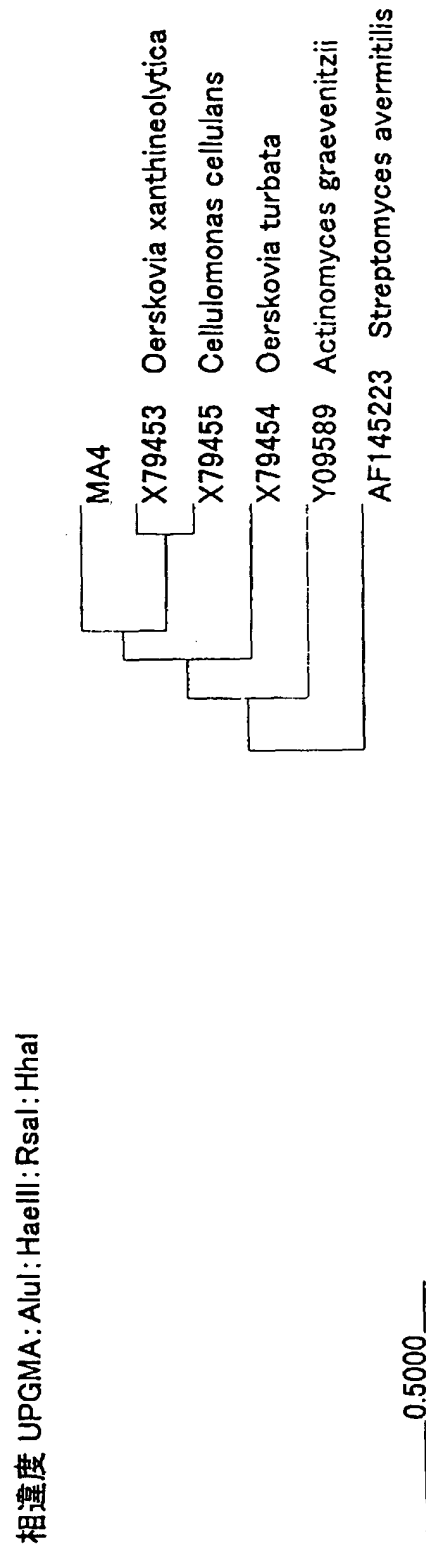


【図 2 6】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



【図 2 7】



【図 28】

分離脱窒菌の 16S rDNA の RFLP (制限酵素 *Hae*III, *Hha*I, *Alu*I, *Rsa*I, *Srf*I) による分類と塩基配列の相同性検索による同定結果の比較

グループ	菌数	RFLP	塩基配列 (%ホモロジー)
I	20	<i>Enterobacteriaceae</i>	<i>Klebsiella</i> 属 (100%)
II	9	<i>Burkholderia</i> 属.	<i>B.vietnamiensis</i> (96.9%, 92.6%, 93.0%, 94.8%, 93.4%), <i>Burkholderia</i> 属 (100%)
III+IV	12	<i>Ralstonia</i> 属	<i>Ralstonia</i> 属 (92.0%, 94.5%, 94.5%), <i>R.paucula</i> (95.1%, 93.8%), <i>R.eutropha</i> (95.6%, 100%, 96.7%)
V	6	<i>Comamonas acidovorans</i>	<i>C.acidovorans</i> (98.2%, 100%)
VI+VIII	32	<i>Pseudomonas</i> 属	<i>P.putida</i> (97.7%, 99.0%, 99.2%), <i>P.fluorescens</i> (95.8%, 99.5%), <i>P.rhodesiae</i> (98.4%)
VII	20	<i>P.putida</i>	<i>P.putida</i> (100%)
IX	8	<i>P.rhodesiae</i>	<i>P.rhodesiae</i> (98.5%, 99.5%)
X	5	<i>P.stutzeri</i>	<i>P.stutzeri</i> (98.0%, 94.6%, 92.0%)
XI	3	<i>Acinetobacter haemolyticus</i>	<i>A.haemolyticus</i> (96.1%)
XII	18	<i>Pseudomonas</i> 属	<i>Pseudomonas</i> 属 (99.5%)
XIII	1	<i>Acivorax delafieldii</i>	<i>Acivorax delafieldii</i> (94.7%)

【書類名】 要約書

【要約】

【課題】 制限酵素による切断片長の実測データを基に、効率的に既知の塩基配列との類縁性を解析するための塩基配列類縁性検索方法および塩基配列類縁性検索システムを提供する。

【解決手段】 理論値計算部 1 1 は、制限酵素による切断位置を表す制限酵素データ 2 1 と既知の遺伝子の塩基配列を表す増幅後塩基配列データ 2 2 b とを基に、切断片の長さを表す切断長理論値データ 2 3 を生成する。切断長実測値データ 2 5 は、制限酵素で切断した試料を用いた電気泳動画像等に基づいた切断長を持っている。比較部 1 3 は、切断長理論値データ 2 3 および切断長実測値データ 2 5 を基に、非加重結合法等の方法により、切断長の分布の類似度を用いたクラスタ分析を行い、解析結果データ 2 7 を生成する。表示部 1 5 は、解析結果を樹形図等によって図形的に表示する。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願 2 0 0 0 - 2 1 5 1 3 4
受付番号	5 0 0 0 0 8 9 6 1 9 7
書類名	特許願
担当官	塩崎 博子 1 6 0 6
作成日	平成 1 2 年 9 月 2 0 日

<認定情報・付加情報>

【特許出願人】

【識別番号】	591224434
【住所又は居所】	熊本県菊池郡西合志町大字須屋 2 4 2 1
【氏名又は名称】	農林水産省九州農業試験場長

【特許出願人】

【識別番号】	500333981
【住所又は居所】	熊本県菊池郡西合志町大字須屋 2 4 2 1 農林水産省九州農業試験場内

【氏名又は名称】	渡邊 克二
----------	-------

【代理人】

申請人

【識別番号】	100064908
【住所又は居所】	東京都新宿区高田馬場 3 丁目 2 3 番 3 号 O R ビル 志賀国際特許事務所

【氏名又は名称】	志賀 正武
----------	-------

【選任した代理人】

【識別番号】	100108578
【住所又は居所】	東京都新宿区高田馬場 3 丁目 2 3 番 3 号 O R ビル 志賀国際特許事務所

【氏名又は名称】	高橋 詔男
----------	-------

【選任した代理人】

【識別番号】	100089037
【住所又は居所】	東京都新宿区高田馬場 3 丁目 2 3 番 3 号 O R ビル 志賀国際特許事務所

【氏名又は名称】	渡邊 隆
----------	------

【選任した代理人】

【識別番号】	100101465
【住所又は居所】	東京都新宿区高田馬場 3 丁目 2 3 番 3 号 O R ビル 志賀国際特許事務所

次頁有

認定・付加情報（続き）

【氏名又は名称】	青山 正和
【選任した代理人】	
【識別番号】	100094400
【住所又は居所】	東京都新宿区高田馬場3丁目23番3号 ORビ ル 志賀国際特許事務所
【氏名又は名称】	鈴木 三義
【選任した代理人】	
【識別番号】	100107836
【住所又は居所】	東京都新宿区高田馬場3丁目23番3号 ORビ ル 志賀国際特許事務所
【氏名又は名称】	西 和哉
【選任した代理人】	
【識別番号】	100108453
【住所又は居所】	東京都新宿区高田馬場3丁目23番3号 ORビ ル 志賀国際特許事務所
【氏名又は名称】	村山 靖彦

【書類名】 手続補正書

【提出日】 平成12年 8月 4日

【あて先】 特許庁長官 殿

【事件の表示】

【出願番号】 特願2000-215134

【補正をする者】

【識別番号】 591224434

【氏名又は名称】 農林水産省九州農業試験場長 小川 奎

【補正をする者】

【識別番号】 500333981

【氏名又は名称】 渡邊 克二

【代理人】

【識別番号】 100064908

【弁理士】

【氏名又は名称】 志賀 正武

【発送番号】 051296

【手続補正 1】

【補正対象書類名】 特許願

【補正対象項目名】 提出物件の目録

【補正方法】 追加

【補正の内容】

【提出物件の目録】

【物件名】 持分契約書 1

特願2000-215134

(B)20001500149



持分契約書

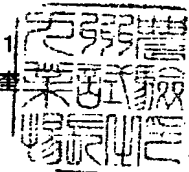
平成12年7月7日

事件の表示 平成12年7月14日付特許願
整理番号 J84516A1

(発明の名称 遺伝子の類縁性検索方法および遺伝子の類縁性検索システム)

上記発明の特許を受ける権利の持分を甲99%、乙1%と定めたことに相違ありません。

(甲) 熊本県菊池郡西合志町大字須屋2421
農林水産省九州農業試験場長 小川 重



(乙) 熊本県菊池郡西合志町大字須屋2421
農林水産省九州農業試験場内 波邊 克二



出 願 人 履 歷 情 報

識別番号 [591224434]

1. 変更年月日 1993年 6月 2日

[変更理由] 名称変更

住 所 熊本県菊池郡西合志町大字須屋2421

氏 名 農林水産省九州農業試験場長

出 願 人 履 歴 情 報

識別番号 [500333981]

1. 変更年月日 2000年 7月14日

[変更理由] 新規登録

住 所 熊本県菊池郡西合志町大字須屋2421 農林水産省九州農業
試験場内

氏 名 渡邊 克二